Available online www.worldscientificnews.com

World Scientific News

An International Scientific Journal

WSN 212 (2026) 128-141 EISSN 2392-2192

A Lightweight Residual Inception Attention Network for Explainable Skin Cancer Classification

Olumba Confidence Chigozirim [1], Olumba Chima Wisdom[2], Olumba Gladys Chinyere[3], Aririguzo Chibueze Favour [4]

[1]Department of Computer Science, Federal College of Education Obudu, Cross River State, Nigeria

[2]Department of Industrial Design Engineering, Chengdu University of Technology, China

[3]Department of Environmental Science and Engineering, Chengdu University of Technology, China

[4]Department of Civil Engineering, Lumkalt Energy Limited, Rukpokwu, Port Harcourt, Rivers State, Nigeria.

*Corresponding Author: olumbaconfidence3@gmail.com

https://doi.org/10.65770/XAMD8420

ABSTRACT

Skin cancer is a common and potentially fatal disease caused by the unnatural growth of skin cells. It can spread to other parts of the body, and early diagnosis significantly affects survival. However, the ability to detect skin cancer early is very difficult. Therefore, in this project, a medical image binary classification system based on deep learning is designed and implemented for distinguishing benign and malignant lesions. An innovative Residual Inception Attention Model is adopted, which uses a multi-branch residual structure and combines depthwise separable convolution, spatially separable convolution, and custom attention mechanism. Interpretability analysis was performed through Grad-CAM heatmap visualization. The training results show that the model has good training results on the test set, which proves that the model has high reliability in medical image diagnosis, can effectively distinguish benign and malignant lesions, and provides strong support for clinical diagnosis.

Keywords: Skin cancer, CNN, TensorFlow/Keras, Attention Mechanism, Grad-CAM, XAI

(Received 10 December 2025; Accepted 16 January 2026; Date of Publication 10 February 2026)

1. INTRODUCTION

The prevalence of skin cancer has been increasing exponentially over the past few decades[1]. Studies have shown that the 5-year survival rate of malignant melanoma is 99% when it is detected in the early stage, and it is significantly reduced to 25% in the advanced stage[2]. But due to the variability of lesion shape, texture, and color, and the similarity between malignant and benign lesions, even experienced dermatologists find it challenging to identify early malignancy from skin images[3] [4].

Methods for automatic diagnosis of skin cancer include the following steps: preprocessing of input images, segmentation of lesion regions, extraction of relevant features, and then classification into malignant or benign types[5]. In recent years, the development of artificial intelligence (AI) and deep learning has provided promising solutions for accurate, scalable, and automated detection[6]. Islam. N et al. pointed out that skin cancer classification based on deep learning has shown good results in early detection and accurate diagnosis[7]. This project aims to design and implement a medical image binary classification system based on deep learning for distinguishing benign and malignant lesions.

As a special type of neural network, Convolutional Neural Networks (CNN) are crucial in deep learning tasks such as image classification and segmentation. This project adopts an innovative Residual Inception Attention network architecture (Residual Inception Attention Model), which uses a multi-branch residual structure and combines depthwise separable convolution, spatially separable convolution and custom attention mechanism. In the data preprocessing stage, a variety of data augmentation techniques, including rotation, translation, scaling, and flipping, are used to improve the generalization ability of the model. The model converts 2D feature maps into 1D feature vectors by global average pooling, which effectively retains global semantic information. The model was trained by binary cross-entropy loss function and Adam optimizer with early stopping strategy and learning rate decay mechanism.

In past studies, although CNNS can accurately classify lesions, it is difficult to understand which features in the image lead to the classification. This makes physicians' trust in the diagnosis challenging and hampers their ability to explain the diagnosis to patients [8]. In this project, interpretability analysis was carried out through Grad-CAM heat map visualization, which confirmed that the model can accurately focus on the lesion area for decision making.

The training results show that the accuracy of the model on the test set is 83.5%, and the AUC value is 94.3%. The training results are good, which proves that the model has high reliability in medical image diagnosis, can effectively distinguish benign and malignant lesions, and provides strong support for clinical diagnosis.

2. RELATED LITERATURE

There are subtle differences in visual features between early skin cancer lesions and benign nevus, so it is easy to miss diagnosis or over-biopsy. In order to improve the accuracy of initial screening, researchers have tried two technical routes, traditional machine learning and deep learning.

Arora et al. [9] constructed the manual feature + support vector machine framework, and input the morphology, color and texture features into the secondary kernel SVM, and achieved 85.7% accuracy and 100% sensitivity on 200 images, but the specificity was only 60%.

This work confirms the clinical feasibility of the "computer-aided + machine learning" route. However, it relies on manual design features, and the end-to-end learning ability is insufficient, and the scalability of SVM for large-scale high-dimensional dermoscopy images is limited.

With the rise of Convolutional Neural Network (CNN), research focuses on deep features. Kiran et al. [10] proposed a hybrid process of "CNN initial classification combined with multi-source features (HOG, LBP, ResNet), feature level fusion, and integrated ML" : Firstly, the lightweight CNN is used to quickly screen, and then the manual and deep descriptors are fused, and finally the decision is made by random forest /XGBoost/AdaBoost voting. On 10,000 Kaggle skin cancer images, the random forest model achieves 96% accuracy (97% precision, 95% recall), outperforming single ANNs or transfer learning. However, this scheme requires phased training and offline feature stitching, has long inference links and large parameters, which is not conducive to real-time deployment of mobile terminals, and lacks explicit focus on key lesion areas, so its interpretability is still limited.

Aiming at the problems of sensitivity and specificity imbalance, large model and weak interpretability of the above works, this study proposes a lightweight "multiple attention residual network". On the one hand, the traditional VGG/ResNet backbone is replaced by a combination of depthwise separable convolution and residual Inception module to achieve multi-scale feature reuse within 1.35 M parameters. On the other hand, parallel channel-spatial attention is embedded in each residual branch so that the network can automatically focus on the key dermoscopic signs such as pigment net and spherical structure without manual features, and the lesion heat map can be directly generated by Grad-CAM, which provides a visual decision-making basis for clinical practice.

3. MATERIAL AND METHOD

3.1. Dataset

The dataset of this study is from Kaggle [11], which is the processed skin cancer images in the ISIC archive. The dataset contains 3297 images of skin cancer, of which 1800 are benign and 1497 are malignant. It provides a solid data foundation for the training of this model.

In order to evaluate the model, the data set was split into three non-overlapping parts: training, validation, and testing, with the proportions of 65%, 15% and 20% respectively (2110 samples in the training set, 527 samples in the validation set, and 660 samples in the testing set), as shown in Figure 1. This division ensures the reliability of the training results.

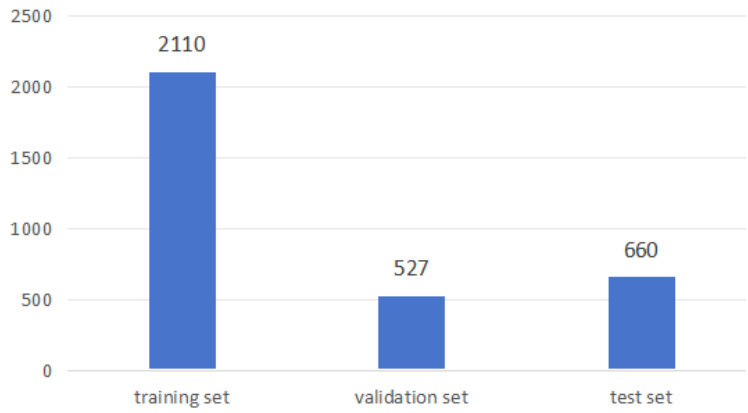

Figure 1. Dataset partitioning.

Four samples are randomly selected from two categories in the data set, as shown in Figure 2, the data meet the requirements and can be used for model training.

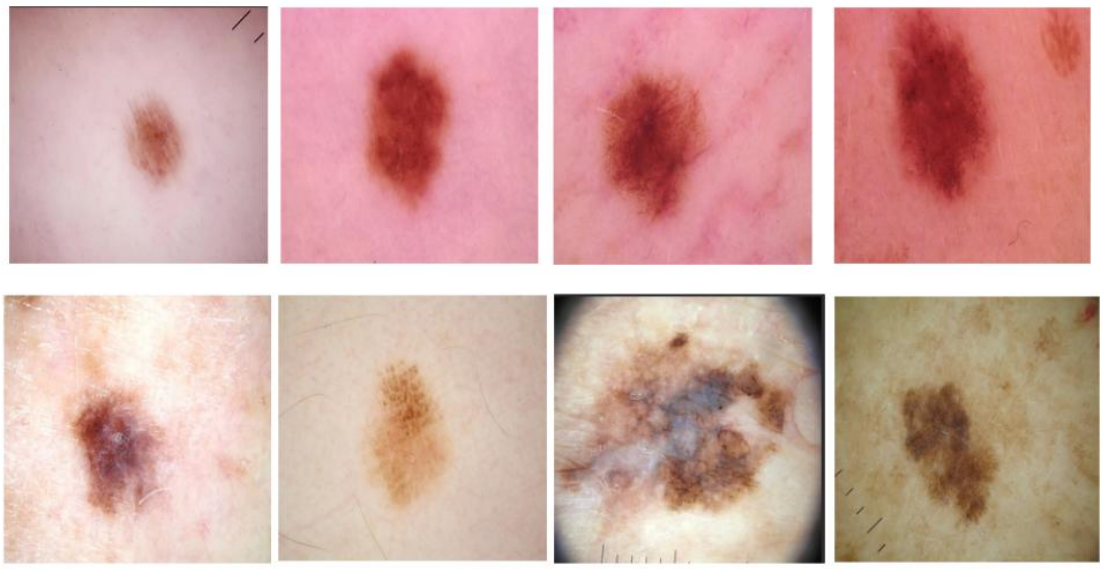

Figure 2. Benign (upper) vs. malignant (lower).

3.2. Proposed Model

This paper presents a lightweight convolutional neural network that integrates attention mechanisms and multi-branch structures, specifically designed for binary classification of skin lesions into malignant or benign.

a) Input Pre-processing

- Images are resized to 224 × 224 to align with pre-trained feature extractors.
- Pixel values are rescaled to [0, 1] for faster convergence.

b) Early Downsampling

- One standard 3×3 conv (stride 2) + MaxPool
- Output: 56×56×32 low-level features

c) Residual-Inception Blocks

Each block contains four parallel paths:

- 1×1 conv
- Depth-wise separable 3×3 conv
- Spatial-wise separable 5×5 conv (split into 1×5 and 5×1)
- AvgPool + 1×1 conv

The four feature maps are concatenated along the channel axis, fused with a 1×1 conv, and fed into a custom Dual-Attention module:

d) Channel Attention:

- GAP & GMP → bottleneck Dense → Sigmoid
- Spatial Attention: 1×1 conv → Sigmoid.
  This adaptively recalibrates features along both channel and spatial dimensions.

e) Progressive Deepening

- Channels: 64 → 128 → 256 → 512
- Spatial resolution: 56×56 → 28×28 → 14×14 → 7×7
- MaxPool inserted between stages.

f) Classification Head

- Global Average Pooling → 512-D vector
- FC 256 + ReLU + Dropout
- FC 128 + ReLU + Dropout
- FC 1 + Sigmoid → probability [0, 1]

g) Training Strategy

- Optimizer: Adam
- Loss: Binary Cross-Entropy
- Callbacks: Early-Stopping, Reduce-LR, Model-Checkpoint

Data Augmentation

Real-time heavy augmentation is applied to dermoscopy images during training:

- Random Rotation: ±0°–40° (p = 0.5)
- Random Translation: ±20 % of image size (horizontal & vertical)
- Random Shear: ±0.2 rad
- Random Zoom: uniform scale in [0.8, 1.2]
- Random Horizontal Flip: p = 0.5

These transformations increase sample diversity, yielding a more generalizable model and reducing overfitting.

3.3. Evaluation Strategy

Accuracy, precision, recall, F1 value, loss value, and Area Under ROC Curve (AUC) were used to evaluate the performance of skin cancer classifiers. The proportion of instances that are correctly classified among all samples is denoted by accuracy. The precision measure is the fraction of accurate positive predictions over all model positive predictions. Recall measures how accurately a classifier identifies successful instances. Precision and recall are balanced by the F1 value, which is the harmonic mean of the two metrics. The specific explanation and calculation methods are as follows:

In this report:

- Tp (True Positive) is the number of malignant lesions correctly classified as malignant.
- False Positive (Fp) is the misclassification of a malignant lesion as a benign one
- Tn (True Negative) means that a benign lesion is correctly classified as benign.
- False Negative (Fn) is the number of benign lesions misclassified as malignant.

(I) Accuracy

Accuracy is the proportion of correctly predicted pairs (positive + negative pairs) over the total number of samples, measuring the proportion of all samples that the model predicts correctly.

$$\text{Accuracy} = \frac{\text{Tp} + \text{Tn}}{\text{Tp} + \text{Fn} + \text{Tn} + \text{Fp}} \quad (1)$$

(II) Precision

Precision is the fraction of positive predictions that are actually positive. It measures how many of the positive predictions of the model can be trusted to avoid false positives.

$$Precision = \frac{\mathrm{Tp}}{\mathrm{Tp} + \mathrm{Fp}} \tag{2}$$

(III) Recall

Recall represents the fraction of samples predicted to be positive out of those that were actually positive. Measures whether the model can cover all true positive samples as much as possible.

$$Recall = \frac{\mathrm{Tp}}{\mathrm{Tp} + \mathrm{Fn}} \tag{3}$$

(IV) F1-Score

The F1-score is the harmonic mean of precision and recall, and it avoids the extreme cases where one measure is too high and the other is too low.

$$F1\text{-}Score = 2\times \frac{Precision \times Recall}{Precision + Recall} = \frac{2\mathrm{Tp}}{2\mathrm{Tp} + \mathrm{Fn} + \mathrm{Fp}} \tag{4}$$

(V) Loss

The loss value represents the metric that the model measures the gap between the predicted result and the true label during the training process, which is the direct goal of model optimization. The smaller the loss value, the smaller the prediction error of the model on the current training data.

This project is a binary classification task, and Binary Cross-Entropy is used to calculate the loss value.

(VI) Area Under ROC Curve (AUC)

AUC is a measure of how well a binary classification model can rank true positives before true negatives. The ROC curve takes false positive rate (FPR) as the horizontal axis and true positive rate (TPR) as the vertical axis, where:

TPR (True Positive Rate) : Recall, the ability to cover true positive samples:

$$\mathrm{Recall} = \frac{\mathrm{Tp}}{\mathrm{Tp} + \mathrm{Fn}} \tag{5}$$

FPR（False Positive Rate）：The proportion of false positives:

$$FPR = \frac{\mathrm{Fp}}{\mathrm{Fp} + \mathrm{Tn}} \tag{6}$$

AUC values range from 0 to 1, with closer to 1 indicating better performance. AUC = 0, the model has no discrimination ability at all, which is equivalent to random guessing. AUC = 1.0: The model perfectly distinguishes all samples; AUC > 0.5 indicates that the model is discriminative; larger values are better.

3.4. Environment Execution

The environment for my project looks like Table3-1:

Table 3-1. Hardware and Software Information.

Type	Manufacturer /Library	Version
Computer	Lenovo	XiaoXinPro 16ACH 2021
CPU	AMD	Ryzen 7 5800H
GPU	NVIDIA	GeForce RTX 3050 Laptop 4G
Memory	Samsung	16GB DDR4 3200MHz (8GB + 8GB)
Operating System	Microsoft Windows	Windows 11 25H2 26200.6725
Environment	Anaconda	Anaconda3
Programming Language	Python	3.8.20
Library	Tensorflow-gpu	2.10.0
Library	keras	2.10.0
Library	numpy	1.22.0
Library	matplotlib	3.7.5
Library	pandas	2.0.3
Library	opencv-python	4.4.0.44

4. EXPERIMENTAL RESULTS

4.1. Performance Results Using the Dataset

According to the above calculation method, the performance of this model is calculated, and the performance in the test set is shown in Table4-1 and Table4-2：

Table 4-1. Classification performance information.

Types	Classification report	
	Benign	Malignant
precision	0.88	0.83
recall	0.86	0.86
f1-score	0.87	0.85
support image	360	300

Table 4-2. Overall performance information.

Tpye	Precision	Accuracy	Loss	Recall	AUC	Tp	Fp	Tn	Fn
Results	0.83	0.86	0.3	0.86	0.94	258	52	308	42

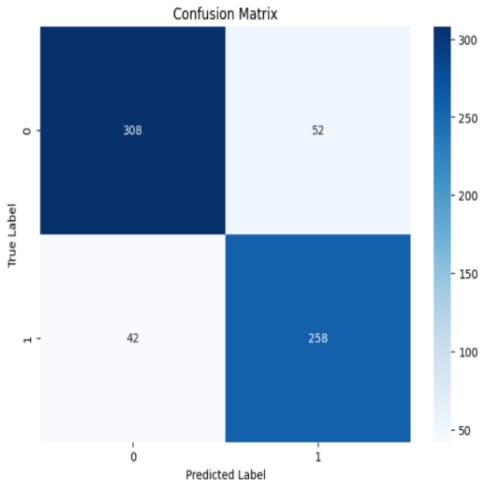

Figure 3. Confusion Matrix.

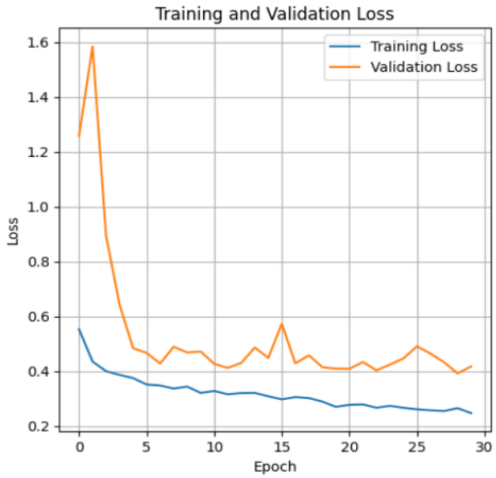

Figure 4. Training and Validation Loss.

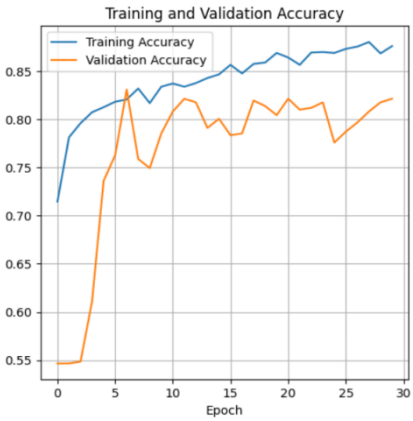

Figure 5. Training and Validation Accuracy.

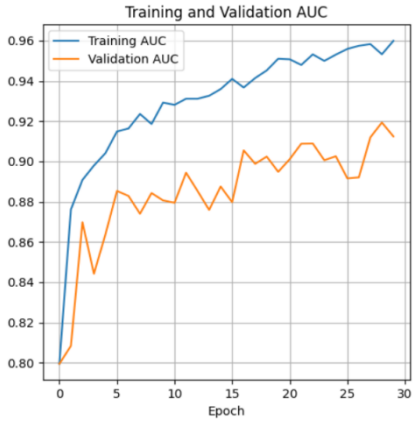

Figure 6. Training and Validation AUC.

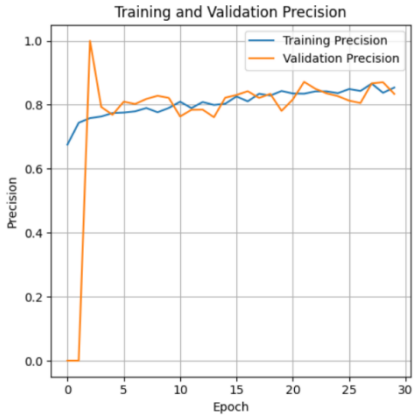

Figure 7. Training and Validation Precision.

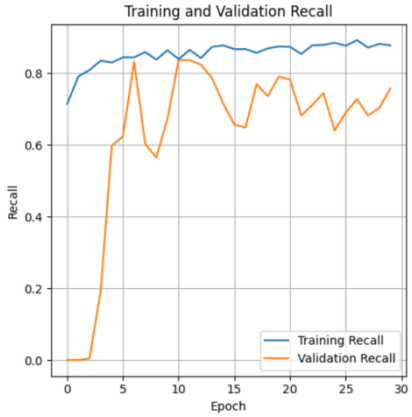

Figure 8. Training and Validation Recall.

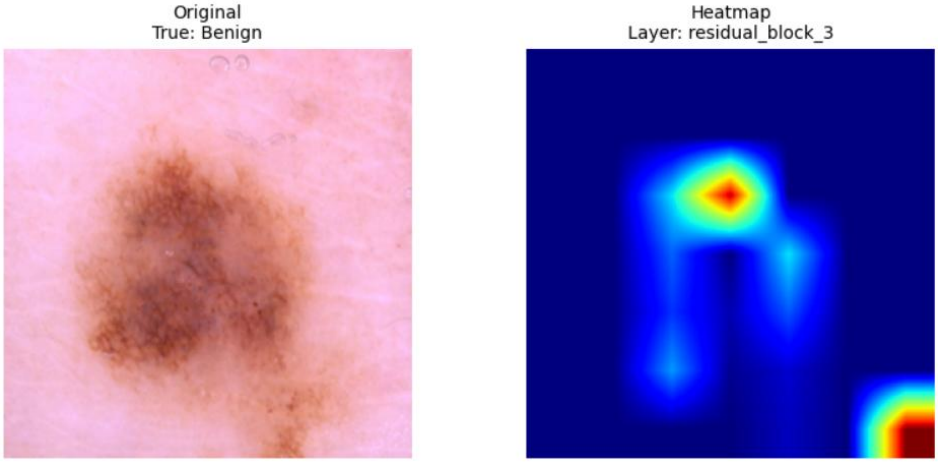

Figure 9. Heatmap

4.2. Discussion

According to the data records of the whole training process in Figure3-9, the training process of the model is divided into three stages.

In Epoch 1-3, it is observed that the training Loss decreases from 0.55 to 0.40, and the validation Loss decreases from 1.26 to 0.89, both of which decrease synchronously and the gap is large, which is analyzed as underfitting state, and the training Accuracy increases from 0.71 to 0.80. However, the verification Accuracy hardly changes. Precision directly dropped to 0 on the validation set, indicating that all samples were judged as benign by the model in the first three epochs and were not correctly identified as malignant, so the Recall was also 0. However, the AUC increased from 0.80 to 0.87, indicating that the criteria are slowly developing.

During epochs 4-10, we observe that the training Loss decreases from 0.40 to 0.32, and the validation Loss decreases from 0.89 to 0.48. Both decrease synchronously and the gap Narrows rapidly, indicating that the model starts to effectively learn task-related features. In the validation set, Precision jumps from 0 to 0.82, and Recall jumps from 0 to 0.67, showing a significant "jump", indicating that the network has initially learned to identify malignant samples, and the classification threshold begins to play a role. AUC increased steadily from 0.87 to 0.88. Validation Accuracy increased from 0.74 to 0.82. The gap between training and validation Accuracy Narrows from 25% to 6%, reflecting the continuous improvement of model generalization ability.

During epochs 11-20, we observe that the training Loss decreases slowly from 0.32 to 0.27, and the validation Loss decreases from 0.48 to 0.41, but the validation curve shows a clear "zigzag" fluctuation: The validation Loss briefly rebounded to 0.49 and 0.46 for epochs 14 and 18, and then fell back, showing typical local minimum oscillation. The verification Accuracy fluctuates between 0.78 and 0.83. The fluctuation of Precision (0.76-0.83) and Recall (0.56-0.84) is more significant, which reflects that the classification threshold is sensitive to the data distribution.

AUC maintains a nearly monotonic upward trend and finally reaches 0.90, indicating that the model's ability to rank samples continues to increase, but the threshold division is affected by fluctuations.

In Epoch 21-30, it is observed that the training Loss continues to decrease to 0.25, and the validation Loss stops decreasing and rises slightly, showing a typical signal of overfitting. The training Accuracy reaches 0.96, while the validation Accuracy stagnates at 0.82. The training AUC was 0.98, and the validation AUC was 0.91, widening the gap to 7%. The fluctuation convergence of Precision (0.83 ± 0.02) and Recall (0.72 ± 0.04) indicates that the prediction behavior of the model tends to be stable.

The final result, Accuracy is 0.86, indicating that the overall classification accuracy is high. The AUC is 0.94, which indicates that the model is very discriminative. The Precision is 0.83, which means that 83% of the samples predicted as malignant are correct. The Recall was 0.86, indicating that 86% of the samples that were actually malignant were successfully detected. The F1-score is 0.85, and the overall performance is good.

4.3. Fair Comparison with Other Deep Learning Models

In order to make a fair comparison, the comparison scope is limited to models trained on the ISIC dataset (1800 benign mole images and 1497 malignant mole images).

The improved multi-attention VGG19 model proposed by Xu et al. [11] adopts the improved VGG19 Mixed Pair Attention (CBAM) mechanism and uses extremely random tree as the classifier.

Table 4-3. Fair Comparison with Other Deep Learning Models.

Type	Improved Multiattention VGG19[11]	Proposed
Accuracy	0.83	0.86
precision	0.83	0.83
Recall	0.83	0.86
F-score	0.83	0.86
AUC	0.81	0.94
parameters	15.1 M	1.35 M

4.4. Comparison with Existing Literature

Mallick et al. [8] proposed a single convolutional lightweight neural Network (CNN) architecture incorporating an attention mechanism, where a "one-parameter" attention layer is inserted after the last convolutional block (11×11×32) to generate a spatial weight map, which is pixel-by-pixel multiplied with the original feature map. Finally, compared with the results of the traditional transfer learning model Inception-v3, the conclusion is that the accuracy of CNN+Attention in the training and testing phase is significantly better than Inception V3, the training accuracy is increased by 11.5 percentage points to 93.5%, and the testing accuracy is increased by 7 percentage points. It reaches 88%, which proves that the attention mechanism effectively improves the diagnostic accuracy by focusing on key features. However, compared with this project, the attention mechanism only has spatial attention, but the input resolution analysis is only 180×180, and the parameter number is 112k, which is more suitable for edge deployment but lacks interpretability.

Maqboo. Z et al. [4] proposed a model VGG16-CASA based on the framework of "Transfer learning + Dual Attention Mechanism (CASA)". VGG16 takes the deep stacked small convolution kernel (3×3) as the core and gradually extracts the low, medium, and high dimensional features of the image through multi-layer convolution. The authors believe that stacking small convolution kernels (3×3) can capture finer lesion textures, and the final test accuracy is 88.62%. However, because there is no residual link to reduce redundant parameters, the number of parameters is as high as 62.6M, the inference speed is about 10ms, and the model training is cumbersome.

Xu et al. [11] proposed an improved multi-attention VGG19 model, which is consistent with VGG16 and is a deep convolution stack, but a hybrid attention module (channel attention + spatial attention) is inserted between the convolutional layers to solve the problem of "uniform processing of features and insufficient focus on key lesion areas" in traditional VGG19. At the same time, Extremely Randomized Trees are used as the classifier instead of the fully connected layer to reduce the parameter redundancy and improve the classification efficiency.

Compared with the above studies, this project designs a lightweight model that integrates depthwise separable convolution, spatially separable convolution, multi-branch residual structure, and hybrid attention mechanism. When the number of parameters is only 1.35M, the model achieves 85.7% accuracy and 94.3% AUC on the test set, which is significantly better than the traditional method of Arora et al., and the number of parameters is much lower than that of large models such as VGG16-CASA while maintaining high accuracy. In addition, the training process visualization and confusion matrix analysis are introduced to enhance the interpretability of the model, which provides an efficient and reliable deep learning solution for the auxiliary diagnosis of skin diseases.

5. CONCLUSION, LIMITATION, FUTURE WORK

This report introduces a new Residual Inception Attention Model for skin cancer binary recognition. The combination of depthwise separable convolution and spatially separable convolution reduces the number of parameters to 1.35M, which is suitable for edge device deployment. At the same time, the dual attention mechanism of channel attention and spatial attention is used to let the model focus on the lesion area. Finally, it shows Accuracy 85.8 %, Precision 83.5 %, Recall 86.0 %, AUC 94.3 % on the test set, which proves that it has high usability and accuracy. At the same time, Grad-CAM is used for visualization to show the area of model decision and enhance the trust of doctors in the results.

However, the limitation of this model is that the test data set is small, only more than three thousand samples. The influence of skin color, hair, blood vessels on model training is not considered, and no corresponding optimization is done. In the future, new large data sets such as HAM10000 can be introduced for strengthening training. Meanwhile, in the future, more models can be added to further optimize the model.

References

[1] A. Mahbod, G. Schaefer, C. Wang, G. Dorffner, R. Ecker, and I. Ellinger, ‘Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification’, *Comput. Methods Programs Biomed.*, vol. 193, p. 105475, Sept. 2020, doi: 10.1016/j.cmpb.2020.105475.

[2] H. K. Koh, ‘Melanoma Screening: Focusing the Public Health Journey’, *Arch. Dermatol.*, vol. 143, no. 1, pp. 101–103, Jan. 2007, doi: 10.1001/archderm.143.1.101.

[3] G. McKnight, J. Shah, and R. Hargest, ‘Physiology of the skin’, *Surg. Oxf.*, vol. 40, no. 1, pp. 8–12, Jan. 2022, doi: 10.1016/j.mpsur.2021.11.005.

[4] Z. Maqbool, S. Pumrin, and N. Panitantum, 'Diagnosis of Skin Cancer via Transfer Learning with Combined Channel attention and Spatial Attention', in *2024 28th International Computer Science and Engineering Conference (ICSEC)*, Nov. 2024, pp. 1–5. doi: 10.1109/ICSEC62781.2024.10770743.

[5] S. Majumder and M. A. Ullah, 'Feature extraction from dermoscopy images for melanoma diagnosis', *SN Appl. Sci.*, vol. 1, no. 7, p. 753, June 2019, doi: 10.1007/s42452-019-0786-8.

[6] S. M. Thwin and H.-S. Park, 'Advancing Skin Cancer Detection: Integrating Attention-Driven Transfer Learning and Autoencoder-Decoder Fusion', in *2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Feb. 2025, pp. 0638–0643. doi: 10.1109/ICAIIC64266.2025.10920870.

[7] N. Islam, J. T. Raya, M. T. Maisha, and D. Md. Farid, 'Feature Fusion with Attention Mechanism for Skin Cancer Classification', in *2023 6th International Conference on Electrical Information and Communication Technology (EICT)*, Dec. 2023, pp. 1–6. doi: 10.1109/EICT61409.2023.10427842.

[8] P. K. Mallick, A. Sharma, N. Padhy, B. Sahoo, and J. Gochhayat, 'Integrating Attention Mechanism for Accurate Skin Cancer Diagnosis', in *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, Feb. 2024, pp. 785–790. doi: 10.1109/ESIC60604.2024.10481565.

[9] G. Arora, A. K. Dubey, Z. A. Jaffery, and A. Rocha, 'Bag of feature and support vector machine based early diagnosis of skin cancer', *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8385–8392, June 2022, doi: 10.1007/s00521-020-05212-y.

[10] T. S. R. Kiran, Y. S. Madhuri, P. Venkata Bala Annapurna, A. Lakshmanarao, and B. S. N. Murthy, 'A Novel Approach to Skin Cancer Detection using CNN, ResNet, and Hybrid Feature Extraction', in *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Oct. 2024, pp. 217–222. doi: 10.1109/ICSSAS64001.2024.10760465.

[11] H. Xu, L. Jin, T. Shen, and F. Huang, 'Skin Cancer Diagnosis based on Improved Multiattention Convolutional Neural Network', in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Mar. 2021, pp. 761–765.

doi:10.1109/IAEAC50856.2021.939072